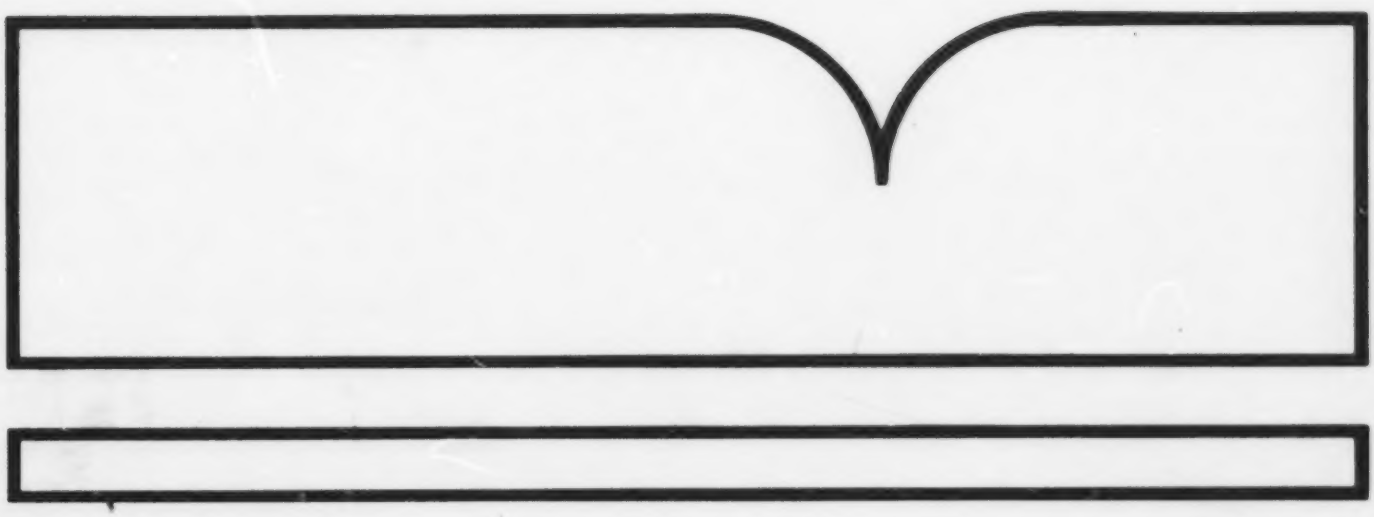
PB81-206013

The Carcinogen Assessment Group's Final Risk Assessment on Arsenic

(U.S.) Environmental Protection Agency
Washington, DC

2 May 81



U.S. Department of Commerce
National Technical Information Service
NTIS

PB81-206013

EPA 600/6-81-002
May 1981

THE CARCINOGEN ASSESSMENT GROUP'S

FINAL RISK ASSESSMENT

ON

ARSENIC

Roy E. Albert, M.D.
Chairman

MAY 2 1980

PARTICIPATING MEMBERS

Elizabeth L. Anderson, Ph.D.
Larry D. Anderson, Ph.D.
Steven Bayard, Ph.D.
Chao W. Chen, Ph.D.
John R. Fowle III, B.S.
Bernard H. Haberman, D.V.M., M.S.
Charalingayya B. Hiremath, Ph.D.
Chang S. Lao, Ph.D.
Robert McGaughy, Ph.D.
Beverly Paigen, Ph.D.
Dharm V. Singh, D.V.M., Ph.D.
Todd W. Thorslund, Sc.D.

REVIEWERS

Bernard Altshuler, Ph.D.
David L. Bayliss, Ph.D.
William J. Blot, Ph.D.

PRODUCT OF:
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

NOTICE

THIS DOCUMENT HAS BEEN REPRODUCED FROM THE BEST COPY FURNISHED US BY THE SPONSORING AGENCY. ALTHOUGH IT IS RECOGNIZED THAT CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED IN THE INTEREST OF MAKING AVAILABLE AS MUCH INFORMATION AS POSSIBLE.

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/6-81-002	Problem-Oriented	PB81 206013

4. TITLE AND SUBTITLE	5. REPORT DATE
Final Risk Assessment on Arsenic	5-2-81
	6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Carcinogen Assessment Group	

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
Carcinogen Assessment Group Office of Health and Environmental Assessment Office of Research and Development U.S. Environmental Protection Agency Washington, D.C. 20460	11. CONTRACT/GRANT NO. N/A

12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Office of Health and Environmental Assessment Office of Research and Development U.S. Environmental Protection Agency Washington, D.C. 20460	Final Report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES
Prepared for the EPA Office of Air Quality Planning and Standards

16. ABSTRACT

An estimate of the potential cancer hazard to populations exposed to airborne arsenic is given. Data analyzed are from occupational studies, estimates of ambient exposure and the number of cancer cases induced in the general or "non-industrial" exposed populations. Using a linear model and data from three studies [Pinto et al. (1977), Ott et al. (1974), and Lee and Fraumeni (1969)], it is estimated that the increase in the lung cancer rate per increase of 1 ug/m^3 of atmospheric arsenic was 9.4%, 17.0% and 3.3% respectively, with a geometric mean of these values calculated to be 8.1%. Using this value and data from a revised 1979 Stanford Research Institute arsenic exposure assessment, lifetime risk of respiratory cancer to those near cotton gins is calculated to be as high as 1.34 x 10^{-2} and higher than 1 x 10^{-5} for about 2.81 million people due to all sources of air arsenic exposure. Total atmospheric arsenic caused deaths due to respiratory cancer is estimated to be within a 95% confidence interval of 19.4 to 1.2 deaths annually.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
NTIS, OAQPS, OHEA - PUBLIC	Unclassified	34 pages
	20. SECURITY CLASS *(This page)*	22. PRICE
	Unclassified	

EPA Form 2220-1 (9-73)

ARSENIC

INTRODUCTION AND SUMMARY

The complete documentation of the arsenic health hazard consists of three parts: 1) a summary of the human health effects (EPA 1980), 2) an estimate of the number of people exposed to arsenic in the ambient air from various identifiable sources and the concentrations to which they may be exposed (SRI 1979), and 3) an estimate of the magnitude of the cancer hazard to the exposed populations. This document is concerned with part three in this series. The purpose is to give a rough estimate of the potential cancer hazard that can be used to guide regulatory decisions that must consider not only the hazard but the costs and benefits associated with implementing actions to reduce the hazard. This analysis utilizes data from occupational epidemiology studies and estimates of ambient exposure to make a rough estimate of the lifetime risk of cancer attributable to those sources and the number of cancer cases induced in the general or "non-industrial" exposed population each year from those sources under the assumption of steady state exposure to the population.

As noted by the EPA arsenic health effects assessment document (EPA 1980), numerous epidemiological studies exist relating increased cancer rates to arsenic exposure. Based on these studies, the Carcinogen Assessment Group concludes that there is substantial evidence that atmospheric arsenic is a human carcinogen. However, many of those studies are inappropriate to use for air arsenic risk assessment since the route of exposure is different, or it is impossible to make a reasonable estimate of the cohort's lifetime average exposure.

Three studies, Lee and Fraumeni (1969), Ott et al. (1974), and Pinto et al. (1977), contained enough pertinent information to make independent quantitative estimates of human cancer risk due to human exposures to atmospheric arsenic.

These studies allow a dose-response relationship to be established between the standard mortality ratio for respiratory cancer and crude estimates of the lifetime average exposure to atmospheric arsenic. The crudeness of the exposure estimates in those studies is due to such factors as high variability in the chemical measurement of arsenic, a scarcity of monitoring data, and the necessity of working from summarized data tables presented in the papers rather than complete data on all individuals.

The EPA arsenic health effects document (EPA 1980) points out additional uncertainties in the studies. However, by accepting the data in spite of its recognized limitations, and making a number of simplifying assumptions concerning dose-response relationships and exposure patterns, it was possible to estimate the carcinogenic potency of arsenic. Using a linear model it was estimated that the increase in the lung cancer rate per increase of 1 ugm/m^3 of atmospheric arsenic was 9.4% (Pinto et al.), 17.0% (Ott et al.), and 3.3% (Lee and Fraumeni). The consistency of these estimates is very good considering the relative crudeness of the data upon which they are based.

The geometric mean of the rate estimates from the three studies was calculated to be 8.1%. Using this value as a best estimate, and the information contained in Table II of the recently revised 1979 SRI arsenic exposure document, it was calculated that exposures to arsenic resulted in a lifetime risk of respiratory cancer as high as 1.34×10^{-2} for a few people exposed in close proximity to cotton gins, and higher than 1×10^{-5} for about 2.81 million people due to all sources of air arsenic exposure. It was also estimated that total atmospheric arsenic caused on the average 4.8 deaths due to respiratory cancer each year, of which about 74% were due to copper smelter operations. The total estimated respiratory cancers due to atmospheric arsenic exposure represent about 0.006% of the U.S. total based on 1973 data.

If the linear model and exposure estimates are correct, so that the only source of variability is due to the estimates of the linear slope parameter, a 95% confidence interval for the above risk estimates may be obtained. Assuming that each study gives an unbiased escimate of the slope parameter and that these estimates are log normally distributed, it follows that upper and lower 95% confidence bounds can be obtained by multiplying the required quantity by about 4 and 0.25 respectively. Thus, 95% statistical confidence bounds would range from 19.4 to 1.2 atmospheric induced respiratory cancer deaths per year. Since it is rather unlikely that population exposure estimates are underestimated by a factor more than two or the true response model falls off slower than the linear model, a non-rigorous absolute upper bound is probably around 40.0 per year. On the other hand, while it is not observable it is plausible that the true response model is non-linear, falling off much faster than the linear model and the exposure estimates on a lifetime basis are considerably over estimated giving a non-rigorous lower bound of virtually zero.

ESTIMATION OF PARAMETERS IN DOSE-RESPONSE MODEL RELATING STANDARD MORTALITY RATIO TO AVERAGE LIFETIME ARSENIC ATMOSPHERIC EXPOSURE

In order to relate lifetime probability of developing respiratory cancer to lifetime average exposure to atmospheric arsenic, we assume that for low exposures the lifetime probability of death due to respiratory cancer can be represented by the linear equation

$$P = \alpha + \beta x$$

where α is the respiratory cancer rate in the absence of atmospheric arsenic, x is the average lifetime exposure to atmospheric arsenic, and the term β is the change in the respiratory cancer rate for each increase of one unit of atmospheric arsenic.

In addition, if we assume that the standard mortality ratio, R, is approximately the same for exposures and observation periods over a segment of a cohort's lifespan as it would be over the entire lifespan, it follows that

$$R = \frac{\alpha + \beta}{\alpha} = 1 + (\beta / \alpha)x$$

or

$$R - 1 = (\beta / \alpha)x.$$

If we define $Y = R - 1$, $\beta^* = (\beta/\alpha)$, then $Y = \beta^* x$ and the term β^* is estimated from different standard mortality ratios corresponding to different exposures by fitting a least squares linear equation through the origin. If additional information relating to the precision of the standard mortality ratio, such as

the number of person-years or expected number of cases, is available, it can be used as a weighting factor, W, so that the best estimate of β^* may be expressed as

$$\hat{\beta}^* = \frac{\Sigma WxY}{\Sigma Wx^2} .$$

A quantity, r*, analogous to the correlation coefficient r, can be computed to give a measure of the goodness of fit of the straight line through the origin. The value r* is the square root of the fraction of the total variability in the Y's explained by the regression equation. This value may be expressed as

$$r^* = \sqrt{1 - \frac{\Sigma W(Y - \hat{\beta}^* x)^2}{\Sigma W(Y - \overline{Y})^2}}$$

where $\overline{Y} = \Sigma WY / \Sigma W$.

Once $\hat{\beta}^*$ is obtained, the quantity of main interest $\beta = \alpha\hat{\beta}^*$ is readily obtained when the value α is specified. The lifetime probability of respiratory cancer, α, in the United States was estimated from the U.S. rates for all sexes and races combined found in the 1973 U.S. vital statistics, utilizing the constant segmented model discussed in Appendix I.

It was estimated that $\alpha = 0.036547$, so that $\beta = 0.036547\hat{\beta}^*$ and the probability at level x of atmospheric-arsenic-caused death due to respiratory cancer is

$$P = 0.036547\hat{\beta}^* x$$

and the expected number of deaths per year of N people exposed at level x is

$$E(D) = 0.036547\hat{\beta}^* xN/71.3$$

where 71.3 is the estimated average U.S. lifespan in years, also derived in Appendix I.

It should be emphasized that this approach is very crude and is utilized mainly because:

(1) It is prudent since it is unlikely to underestimate the true risk.

(2) Enough data exists in the published papers to obtain slope estimates. In contrast, the assumption of a non-linear model would require individual dose estimates such as is obtained from radiation monitoring badges which simply are not available for arsenic exposures.

In the next section the terms β^* are developed for each epidemiological study. In the final section the values P and E(D) are obtained for the different sources of atmospheric arsenic emissions in the United States.

EPIDEMIOLOGICAL STUDIES WITH PERTINENT EXPOSURE INFORMATION TO ATMOSPHERIC ARSENIC

Lee and Fraumeni (1969)

Smelter workers exposed to arsenic trioxide and sulfur dioxide in combination showed a definite positive dose-response trend. However, in order to relate the exposure to the average lifetime ugm/m^3 of arsenic in the air breathed, it was necessary to make a number of simplifying assumptions.

First, the average age at death, or the age at the end of the observation period, and the number of years exposed had to be estimated for each of the five cohorts in the study. These estimates and the rationale for their computation are shown in Table 1.

Lee and Fraumeni (Table 5) estimated the observed and expected respiratory cancer cases for three classifications of exposure, "heavy, medium, light," for cohorts 1, 2, and 3-5 combined. Subsequent to the publication of their paper, workplace measurements of arsenic were made corresponding to these

classifications and were published in a 1975 NIOSH document. The average exposures were found to be:

$$\text{heavy} = 1.47 \text{ mg As/m}^3$$

$$\text{medium} = 1.56 \text{ mg As/m}^3$$

$$\text{light} = 0.70 \text{ mg As/m}^3$$

Since the Lee and Fraumeni intuitive classification of "heavy" and "medium" exposure values, based upon actual chemical measurements, turned out to be virtually identical, we have combined them into heavy + medium and light, and lifetime average exposures are given by multiplying the fraction of lifetime exposed times the workplace exposure. The heavy + medium exposure is taken as the average $(1.47 + 1.56)/2 = 1.52$, and the exposure in cohorts 3-5 is taken as the people-years weighted average for the three cohorts. This results in the exposure and SMR shown in Table 2. In Table 3 the calculations of the slope and correlation coefficient are shown. The goodness of fit of the data is illustrated in Figure 1.

Pinto et al. (1977)

In this study a group of 527 retired copper smelter workers were exposed to airborne arsenic trioxide. These employees had an "exposure index" associated with their employment history that corresponded to the product of the number of years worked at a job location times the average urinary arsenic concentration of workers working in the same locations in 1973.

Under the following set of assumptions this "exposure index," I, was converted to a lifetime average exposure of $\mu gm/m^3$ of atmospheric arsenic.

1. The exposure index, I, was multiplied by 0.304 to convert urinary arsenic to airborne arsenic based on the regression equation developed by Pinto et al. (1976).

2. To convert to a 24-hour, 365-day-per-year equivalent exposure, the index was multiplied by the fraction of year actually on the job site, assumed to be 0.21.

3. It was noted that, due to past measurement of airborne arsenic in the 1930's and early 1940's, the actual levels of exposure at that time were 5 to 10 times higher than in 1973. To correct for this fact, it was assumed that the exposure from 1910 to 1942 was 1/2(5 + 10) = 7.5 times as high as in 1973. From 1942 to 1973 the exposure decreased linearly. Under these assumptions we obtain a time-weighted average of an exposure of 5.9 times that observed in 1973. The index is multiplied by this 5.9 factor to correct for higher past exposures.

4. The average age of the workers is taken as 65 + 13.2 = 78.2 years, where 65 represents the fact that the employee had to be retired to be in the sample and 13.2 years represents the average lifespan for white males, given they are alive at age 65. Since about 62% of the sample have died, this is felt to be an acceptable estimate. The index is divided by 78.2 to put exposure on a lifetime average basis.

The exposure in lifetime average $\mu gm/m^3$ from the air is thus:

$$I^* = I \times 0.304 \times 0.21 \times 5.9/78.2 = 4.817 \times 10^{-3} I.$$

Utilizing the information in Pinto et al. (1977, Table 2), we derive new exposures and SMR's shown in Table 4, and calculate the regression and correlation coefficients from this data. In Figure 2, the goodness of fit of the data is illustrated.

It was calculated that $\hat{\beta} = 0.0935$, so that an increase of 9.35% of the general respiratory cancer rate could be expected for each increase of 1 $\mu gm/m^3$ of arsenic in the atmosphere.

Ott et al. (1974)

Workers engaged in the production of insecticides containing lead arsenate and calcium arsenate showed a marked elevation in respiratory cancer rates based upon a retrospective proportionate mortality study.

Due to a well-documented working history and long-standing work place monitoring for airborne arsenic, it was possible to estimate the total absorbed dose of arsenic on a mg basis for each individual. Individuals were grouped into exposure intervals and the mean of the individuals' values in each interval is given in column 1 of Table 5. However, these values were based upon a highly conservative assumption of a 4 m^3 tidal volume in an 8-hour day. If we make the more common assumption of a 10 m^3 tidal volume for a heavy working day, these total mg estimates would be increased by a factor of 2.5. The $\mu gm/m^3$ in the air required to obtain 2.5 times the number of mg in column 1 is calculated assuming a 20 m^3 daily tidal volume and a 71.3-year lifespan in order to equate the exposure in mgs to a lifetime average air exposure. This is done by multiplying by the factor

$$f = \frac{2.5 \times 10^{+3}}{20 \times 365 \times 71.3} = 4.83 \times 10^{-3}$$

Exposure estimates obtained in this manner are shown in column 2 of Table 5. We note that for the highest exposure group only a marginal increase in the SMR is observed.

Many possible explanations exist as to why that highest point does not follow a line through the origin even though that may be the true relationship for lower exposure levels. For example, the high levels of arsenic may have set up competing mortality that caused death from arsenic due to other causes before death due to respiratory cancer could manifest itself, or saturation to arsenic

exposure at the site of action may have occurred. However, the plausibility of these arguments are somewhat suspect since the exposure very likely is not high enough to bring these mechanisms into play. A more likely explanation is that the exposures were calculated under the assumption that respirators were not used but it is known that there was a tendency for some of those most highly exposed to use respirators. This high exposure group gives good evidence that an arsenic-induced respiratory cancer effect is being elicited, but due to the respirator use bias that is not adjustable for, it cannot be used to fit a dose-response relationship to. We proceed by fitting an equation through the origin with the highest exposed group eliminated. We obtain estimates of $\beta^* = 0.1705$, $r^* = 0.8798$. The calculations and graphical presentation of these results are shown in Table 5 and Figure 3.

RISK TO U.S. POPULATION FROM DIFFERENT SOURCES RELEASING ARSENIC INTO THE ATMOSPHERE

The SRI, in a 1979 document, estimated the human exposure to arsenic from selected emission sources. These data were summarized in Table II-1 of that document, which listed exposure intervals from different emission sources and the number of people exposed from each source in each exposure interval.

Using the estimate $\beta^* = \sqrt[3]{(0.0333)(0.0935)(0.1704)} = 0.0808$ which is the geometric mean of the three studies, we have that the lifetime probability of respiratory cancer is estimated to be

$$P \simeq (0.036547)(0.0808)x = 2.953 \times 10^{-3}x$$

and the yearly expected number of respiratory cancer deaths is

$$E(D) \simeq [(0.035647)(0.0808) \div 71.3]xN = 4.14 \times 10^{-5}\ xN$$

where x is the lifetime average exposure to atmospheric arsenic in $\mu gm/m^3$ and N is the total number of people exposed at level x.

Applying these equations to the data in SRI Table II-1, where x is the midpoint of the exposure interval, we obtain the data shown in Table 6. It is interesting to note that the entire exposed population of 2.81 million people have a probability of response greater than 10^{-5}. However, we would expect a total of only 4.8 deaths due to arsenic-caused respiratory cancer to occur each year. Of this number, it is estimated that 3.5, or 74%, would be caused by copper smelters.

RELIABILITY OF ESTIMATES

It is obviously important to have some feeling for the precision of the obtained estimates. A rigorous assessment of a confidence interval for the estimate is precluded by the complexity of the quantity we are attempting to measure and the unknown reliability of the individual elements upon which the

estimates are based. However, under the following assumptions:

1. The response in truly linear.
2. The estimated values of β from the different studies are unbiased and log normally distributed.
3. The exposure estimates are exact.

A 95% confidence interval for the log total number of tumors T, may be expressed as

$$\log T \pm t_{n-1} \sqrt{s^2/n}$$

where $s^2 = \sum_{j=1}^{n} (\log \beta j - \overline{\log \beta})^2/n-1 = 0.1279$ is an estimate of the variance of the log slopes, n = 3 is the number of independent estimates of β and t_{n-1} is the student "t" statistic with n-1 degrees of freedom.

Under these assumptions, a 95% confidence interval for the log T is

$$\log(4.79) \pm 2.94 \sqrt{\frac{0.1279}{3}} = 1.2874 \text{ to } 0.0719$$

and the 95% confidence interval for total lung tumors is 19.38 to 1.18. The actual exposure estimates are undoubtably in error. For various assumed percentage errors of the total exposures, upper and lower bounds are calculated and shown in Table 7.

TABLE 1. DERIVATION OF FRACTION OF LIFETIME EXPOSED TO ARSENIC EMISSION IN WORK PLACE
LEE AND FRAUMENI 1969

Cohort Number	Average age at death or end of observation period	Average years of exposure	Fraction of lifetime exposed[f]
1	26[a] + 15[b] (15,799/947)[c] = 57.7	15[b] + (15,799/947)[c] = 31.7	0.115
2	26[a] + 1/2(15)[d] + (17,460/1,644)[c] = 44.1	1/2(15)[d] + (17,460/1,644)[c] = 18.1	0.086
3	26[a] + (13,108/883)[c] = 40.8	1/2(15 + 10)[e] = 12.5	0.064
4	26[a] + (21,136/1,298)[c] = 42.3	1/2(5 + 10)[e] = 7.5	0.037
5	26[a] + (57,404/3,275)[c] = 43.5	1/2(1 + 5)[e] = 3.0	0.015

Source: Lee and Fraumeni, Table 2.

[a]Assumed average age at start of employment, 26 years, as measured in cohorts 1 and 2.

[b]Given 15 years or more worked prior to 1938.

[c]Average years lived after 1938 = person-years of observation divided by number of people in cohort. Assumes remained employed during observation period.

[d]Assumes 1/2 exposed years prior to 1938 and 1/2 after 1938.

[e]Midpoint of interval of number of years exposed.

[f]Number of years of exposure ÷ average age at death or at end of the obseration period x (0.21 = 40 x 46/(365 x 24) = fraction of year on job site).

TABLE 2. LIFETIME AVERAGE EXPOSURES AND STANDARD MORTALITY RATIOS
LEE AND FRAUMENI 1969

Cohort	Exposure group	Estimated lifetime average exposure in ugm/m^3	Standard Mortality Ratio (SMR) = R
1	H + M	174.8	6.98
	L	80.5	2.50
2	H + M	130.7	5.81
	L	60.2	3.10
3-5	H + M	41.0	2.98
	L	19.0	2.14

TABLE 3

DATA UTILIZED AND COMPUTATIONS MADE TO ESTIMATE REGRESSION AND CORRELATION COEFFICIENT FROM LEE AND FRAUMENI 1969

$Y = R - 1$	W = Expected Number of Cases	x	$\hat{Y} = \hat{\beta}^* x$
5.98	4.3	174.8	5.82
1.50	5.6	80.5	2.68
4.81	3.1	130.7	4.35
2.10	2.9	60.2	2.00
1.98	4.7	41.0	1.37
1.14	10.3	19.0	.63

$$\bar{Y} = \frac{\Sigma WY}{\Sigma W} = \frac{76.014}{30.9} = 2.46$$

$$\beta/\alpha = \hat{\beta}^* = \frac{\Sigma WxY}{\Sigma Wx^2} = \frac{8,091.14}{242,760.51} = .0333$$

$$\Sigma W(Y - \hat{\beta}^* x)^2 = 13.0204$$

$$\Sigma W(Y - \bar{Y})^2 = 94.9649$$

$$r^* = \sqrt{1 - \frac{13.0204}{94,9649}} = .9289$$

TABLE 4

DATA UTILIZED AND COMPUTATIONS MADE TO ESTIMATE REGRESSION AND CORRELATION COEFFICIENT FROM PINTO (1977)

$Y = R - 1$	I	$x = I^*$	$\hat{Y} = \hat{\beta}^* x$
.11	1,514	7.29	.68
.90	2,513	12.11	1.13
1.82	4,317	20.80	1.94
2.04	7,473	36.00	3.37
4.71	10,135	48.82	4.56
7.33	14,712	70.87	6.63

$$\overline{Y} = \frac{16.91}{6} = 2.82$$

$$\beta/\alpha = \hat{\beta}^* = \frac{\Sigma xY}{\Sigma x^2} = \frac{872.38}{9,333.60} = .0935$$

$$\Sigma(Y - \hat{\beta}^* x)^2 = 2.6766$$

$$\Sigma(Y - \overline{Y})^2 = 36.5511$$

$$r^* = \sqrt{1 - \frac{2.6766}{36.5511}} = .9627$$

TABLE 5. LIFETIME AVERAGE EXPOSURES AND STANDARD MORTALITY RATIOS BASED ON DECEASED COHORT
OTT 1974

Exposure mg	ugm/m^3 in Air x	Respiratory cancer mortality ratios	Y = R - 1	$\hat{Y} = \hat{\beta}^*x$
42	0.2	0.6	-0.4	0.03
127	0.6	2.0	1.0	0.10
252	1.2	2.9	1.9	0.20
420	2.0	2.2	1.2	0.34
796	3.9	1.8	0.8	0.66
1,556	7.6	2.1	1.1	1.30
3,533	17.2	3.9	2.9	2.93
6,503	31.6	6.3	5.3	5.38
29,733*	144.3	7.0	5.9	

*omitting highest exposure

$\Sigma XY = 234.04$

$\Sigma X^2 = 1373.21$

$\Sigma Y = 13.8$

$\bar{Y} = 1.725$

$\Sigma Y^2 = 44.56$

$\Sigma(Y - \bar{Y})^2 = 20.76$

$\Sigma(Y - \hat{\beta}^*x)^2 = 4.69$

$$\hat{\beta}^* = \frac{234.04}{1,373.21} = 0.1704$$

$$r^* = \sqrt{1 - \frac{4.69}{20.76}} = 0.8798$$

TABLE 6. EXPECTED LIFETIME PROBABILITY AND NUMBER OF LUNG CANCER DEATHS PER YEAR DUE TO DIFFERENT SOURCES OF ATMOSPHERIC ARSENIC EXPOSURE

Average annual concentration arsenic[a] ugm/m^3 midpoint exposure interval	Lifetime probability of respiratory cancer due to arsenic	Copper smelters	Lead smelters	Zinc smelters	Cotton gins	Pesticide manufacture	Glass manufacture	Total
4.5	1.34×10^{-2}				3.73×10^{-4} (2)			
2.0	5.96×10^{-3}				2.48×10^{-3} (30)			
0.8	2.38×10^{-3}	8.09×10^{-1} (24,420)			2.32×10^{-3} (70)			
0.45	1.34×10^{-3}	3.64×10^{-1} (19,380)			4.28×10^{-3} (230)			
0.20	5.96×10^{-4}	1.15 (137,450)	6.72×10^{-3} (800)	7.56×10^{-2} (9,000)	5.46×10^{-3} (660)		1.08×10^{-2} (1,440)	
0.08	2.38×10^{-4}	4.00×10^{-1} (119,860)	8.74×10^{-3} (2,600)	6.05×10^{-2} (18,000)	3.28×10^{-3} (990)		3.39×10^{-1} (10,140)	
0.045	1.34×10^{-4}	5.38×10^{-1} (286,560)	7.18×10^{-2} (5,100)	1.91×10^{-1} (101,000)	3.54×10^{-3} (1,900)		1.42×10^{-1} (75,180)	
0.020	5.96×10^{-5}	2.61×10^{-1} (313,380)	3.19×10^{-2} (38,000)	1.43×10^{-1} (170,000)	5.46×10^{-3} (6,600)		1.77×10^{-1} (212,040)	
0.0075	2.23×10^{-5}	2.75×10^{-3} (8,810)	1.45×10^{-2} (46,000)	3.47×10^{-2} (110,000)	5.58×10^{-3} (18,000)	1.18×10^{-4} (380)	1.06×10^{-1} (363,870)	
0.0040	1.19×10^{-5}	2.17×10^{-3} (12,960)	1.13×10^{-2} (67,000)	2.26×10^{-3} (13,500)	7.29×10^{-3} (44,000)	1.82×10^{-4} (1,100)	9.66×10^{-2} (583,860)	
Total		3.53	0.15	0.50	0.04	0.00	0.57	4.79

*Note: values in ( ) are numbers of people exposed.

TABLE 7. UPPER AND LOWER BOUNDS FOR TOTAL RESPIRATORY CANCER DEATHS PER YEAR DUE TO ATMOSPHERIC ARSENIC EXPOSURE BASED ON DIFFERENT ASSUMED LEVELS OF THE PRECISION OF THE EXPOSURE ESTIMATES

Assumed level of precision of exposure estimate % deviation from true estimate	Lower Bound	Upper Bound
0%	1.2	19.4
± 50%	0.8	29.1
± 100%	0.6	38.8
± 200%	0.4	58.2

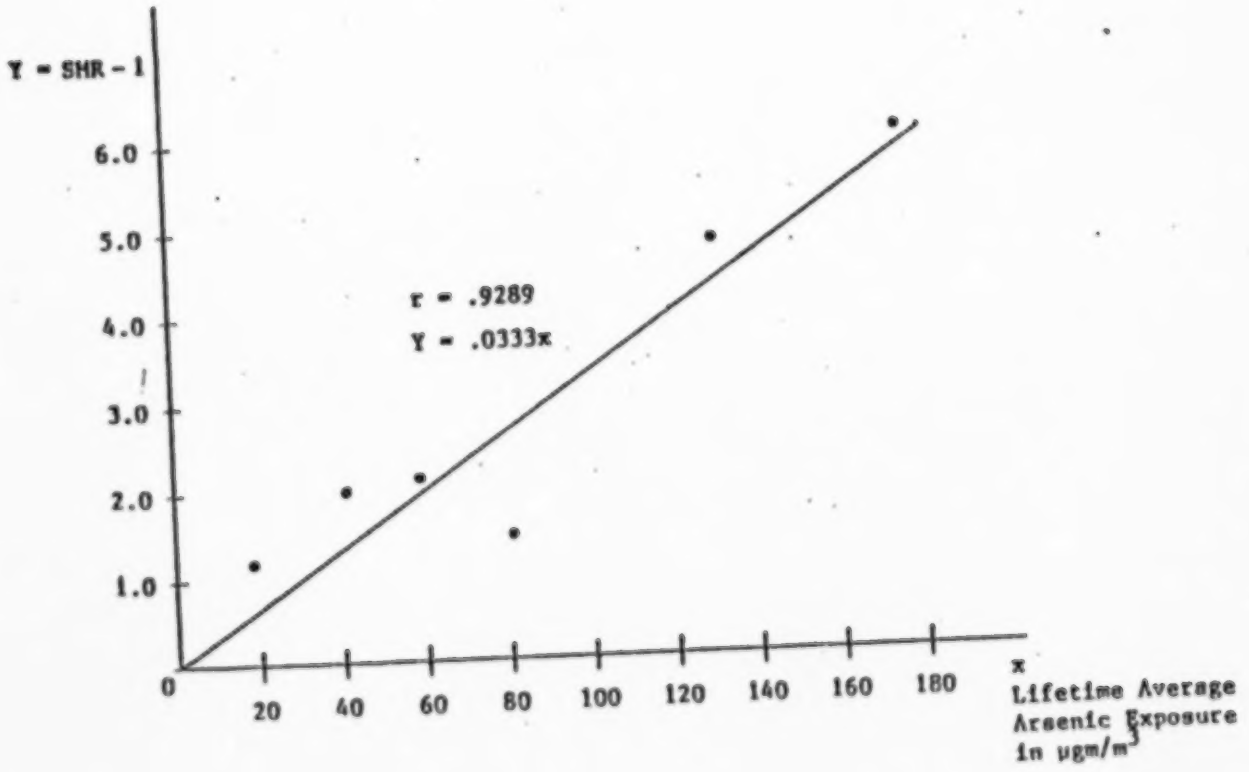


Figure 1. Lee and Fraumeni (1969)

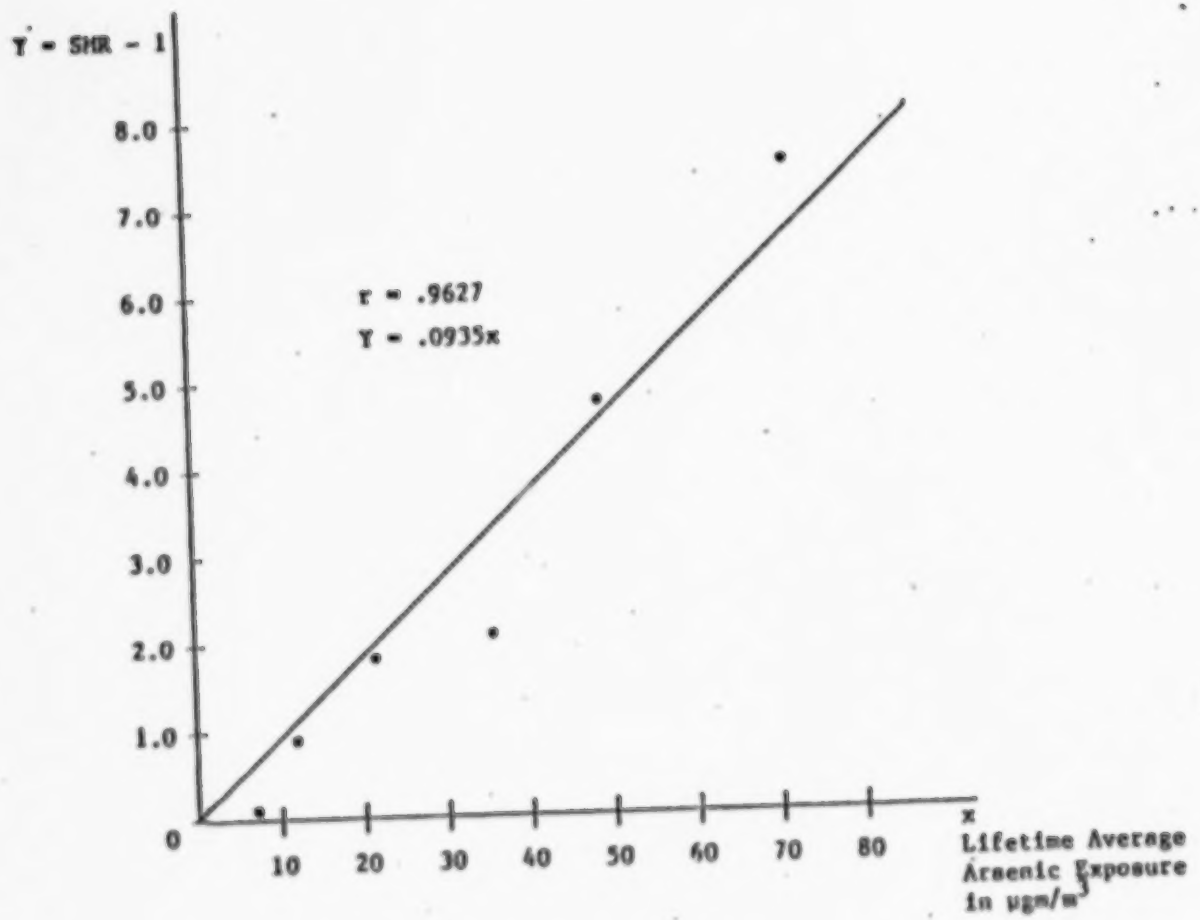


Figure 2. Finto et al. (1977)

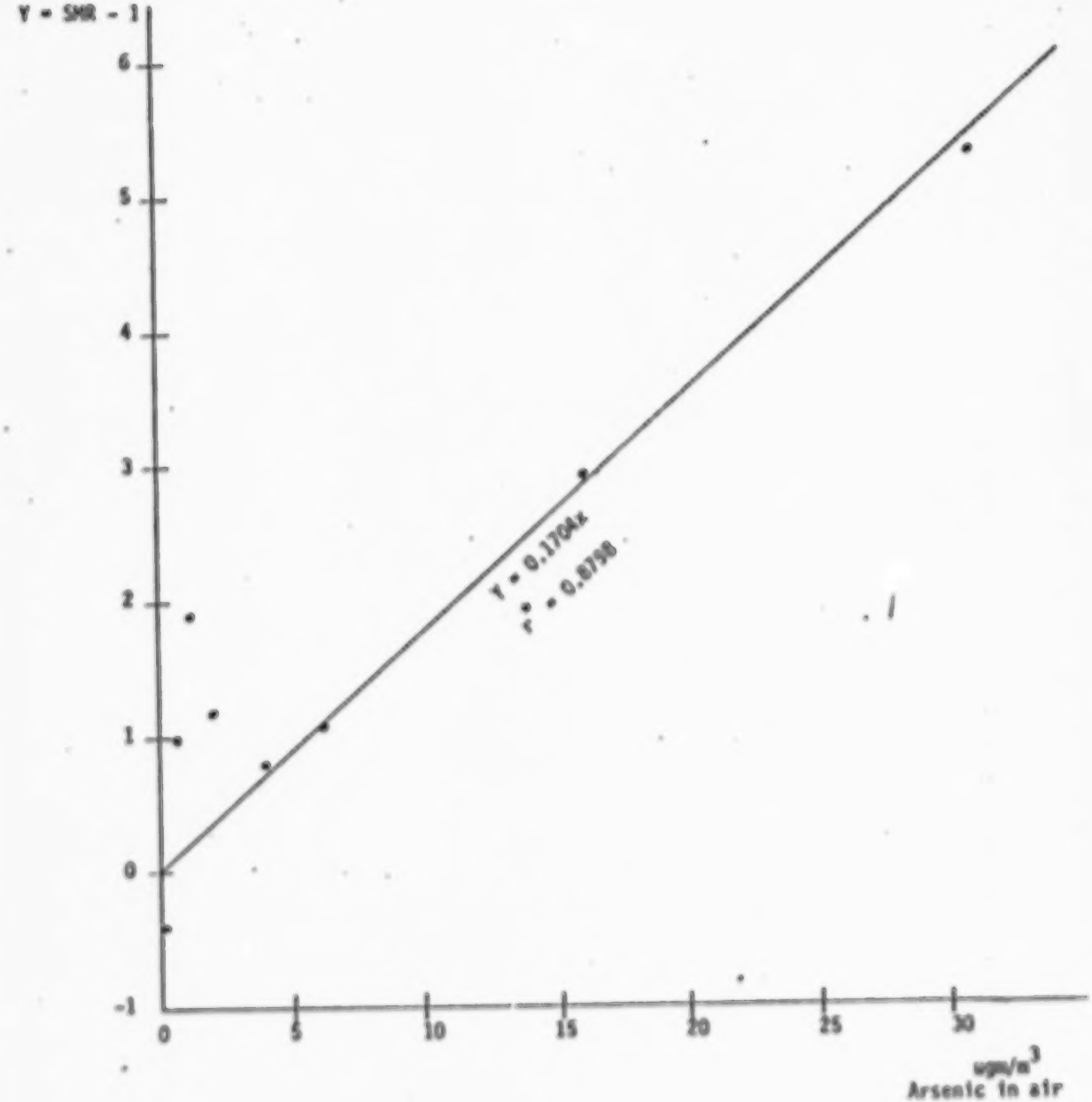


Figure 3. Ott et al. (1974)

REFERENCES

Environmental Protection Agency. December 1979. An assessment of the health effects of arsenic germane to low-level exposure. Revised external review draft.

Lee, A.M., and J.F. Fraumeni. 1969. Arsenic and respiratory cancer in man: An occupational study. J. Natl. Cancer Inst. 42(6):1045-1052.

NIOSH. 1975. Occupational exposure to inorganic arsenic. HEW Publications 75:149.

Ott, M.G., B.J. Holder, and L.H. Gordon. Respiratory cancer and occupational exposure to arsenicals. Arch. Environ. Health 29:250-255.

Pinto, S.S., M.O. Varner, K.W. Nelson, A.L. Labbe, and L.D. White. 1976. Arsenic trioxide absorption and excretion in industry. J. Occupational Medicine 18:677-680.

Pinto, S.S., P.E. Enterline, V. Henderson, and M.O. Varner. 1977. Mortality experience in relation to a measure arsenic trioxide exposure. Environ. Health Perspectives 19:127-130.

SRI-1978, Suta, B.E. 1979. Human exposures to atmospheric arsenic report to EPA.

APPENDIX I

ESTIMATION OF LIFETIME RISK OF DEATH AND YEARS OF LIFE LOST DUE TO A SPECIFIC CAUSE

General Approach

The estimation of the lifetime probability of a disease in the presence of competing causes of death and the resulting life-shorting of the disease is a problem that has received considerable attention. Chiang (1968, pp. 242-68) has given a general solution to the problem using standard methods in competing risk analysis. Gail (1975), using these methods, derives explicit "figures of merit" for measuring the benefit of reduced exposure to environmental carcinogens.

Following Gail, we define the following functions:

$$S_1(t) = P\{\text{survive to} \geq t/c_1 \text{ is only cause of death}\}$$
$$S_2(t) = P\{\text{survive to} \geq t/c_2 \text{ is only cause of death}\},$$

where c_1 and c_2 make up all causes of death. The hazard function or age-specific or instantaneous death rate is denoted as $h_1(t)$, $h_2(t)$, respectively, for causes of death c_1 and c_2. Under the usual assumptions of independence of c_1 and c_2 the total probability of survival until time t and the total hazard from both causes are: $S(t) = S_1(t)S_2(t)$, $h(t) = h_1(t) + h_2(t)$, respectively and

$$S(t) = e^{-\int_0^t h(u)du}.$$

It can be readily shown that the lifetime probability of dying of cause c_2 is

$$Q_2(\infty) = \int_0^\infty h_2(t)S(t)dt$$

and the expected survival time or lifetime is

$$E = \int_0^\infty S(t)h(t)tdt.$$

Thus, given the specific functions $h_1(t)$, $h_2(t)$, the quantities of interest $Q_2(\infty)$ and E can be readily derived.

For two common assumptions concerning the form of $h_1(t)$, $h_2(t)$, the terms $Q_2(\infty)$, E are obtained in the next two sections.

Constant Segmented Model

Often the hazard or age-specific death rate is assumed to be constant over different time intervals but to change at interval boundaries. Formally, let us assume the entire life span is broken up into m mutually exclusive segments or intervals and the hazard functions are defined as follows:

Time Interval	$h_1(t)$	$h_2(t)$	$h(t)$
$0 - t_1$	α_{11}	α_{12}	α_1
$t_1 - t_2$	α_{21}	α_{22}	α_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t_{j-1} - t_j$	α_{j1}	α_{j2}	α_j
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t_{m-1} - \infty$	α_{m1}	α_{m2}	α_m

This is the situation where age-specific rates are estimated from narrow age intervals of typically 1 to 10 years.

In this case, the function $S(t)$, $t_{j-1} \le t \le t_j$, is

$$S(t) = S(t_{j-1}) \cdot e^{-\int_0^{t-t_{j-1}} \alpha_j du} = e^{-\left[(t-t_{j-1})\alpha_j + \sum_{k=1}^{j-1} (t_k - t_{k-1})\alpha_k\right]}$$

so that

$$Q_2(\infty) = \sum_{j=1}^{m} S(t_{j-1}) \int_{t_{j-1}}^{t_j} \alpha_{2j} e^{-(t-t_{j-1})\alpha_j} dt$$

$$= \sum_{j=1}^{m} S(t_{j-1}) \left[1 - e^{-(t_j - t_{j-1})\alpha_j}\right] \alpha_{2j}/\alpha_j \, .$$

This lifetime probability may be viewed intuitively as the sum of the probability of m mutually independent events, namely of dying of the specified disease in the specified time interval. Each of these m probabilities is the product of three probabilities:

1. Probability of death due to cause c_2 given that death occurred in the j^{th} time interval:

$$P_{1j} = \alpha_{2j}/\alpha_j \, ;$$

2. Probability of death between time t_j and t_{j-1} given survival up to time t_{j-1}:

$$P_{2j} = 1 - e^{-(t_j - t_{j-1})\alpha_j} \, ; \text{ and}$$

3. Probability of survival until time t_{j-1}:

$$P_{3j} = S(t_{j-1}) = \prod_{k=1}^{j-1}(1-P_{2k}) = e^{-\sum_{k=1}^{j-1}(t_k - t_{k-1})\alpha_k} .$$

By definition it follows that $P_{31} = 1$, $P_{2m} = 1$, and

$$Q_2(\infty) = \sum_{j=1}^{m} P_{1j}P_{2j}P_{3j} .$$

The expected life span under the segmented model may be expressed as

$$E = \sum_{j=1}^{m} S(t_{j-1}) \int_{t_{j-1}}^{t_j} \alpha_j e^{-(t-t_{j-1})\alpha_j} (t-t_{j-1})dt$$

$$= \sum_{j=1}^{m} S(t_{j-1}) \left(1 - e^{-(t_j-t_{j-1})\alpha_j}\right)/\alpha_j = \sum_{j=1}^{m} P_{2j}P_{3j}/\alpha_j$$

$$= \sum_{j=1}^{m} P_{1j}P_{2j}P_{3j}/\alpha_{2j} .$$

As an example of the utilization of this method, the lifetime probability of death due to respiratory cancer for a randomly selected person in the United States is computed based upon 1973 Vital Statistics. An implicit assumption in utilizing the data in this manner is that the age-specific rates in 1973 will remain constant over time.

From pages 1-8, 1-9, ref. 4, the age-specific death rates for all causes are given and listed under the column αj in Table II-1. The age-specific death rates for malignant neoplasms of respiratory system ICD 160-163 are taken from the same source, pages 1-10, 1-11, and listed

under α_{2j} in Table II-1. From this data the terms P_{1j}, P_{2j}, P_{3j}, and their product Q_{2j} are computed according to the given formulas, the details of which are shown in Table II-1. The end result is an estimated lifetime probability of death due to respiratory cancer of $Q_2(\infty) = .03654667$ and an expected life span of 71.328888 years.

Weibull Model

Doll (1972) and others have pointed out that the Weibull distribution often fits time- and dose-dependent carcinogen response data. One form of a generalized Weibull model is to assume that

$$h_2(t) = \gamma(\alpha + \beta x^m) t^{\gamma-1} ,$$

where t is the age of the individual, x is the level of exposure to a carcinogen and γ, α, β, m are unknown parameters to be estimated from the data. If we make the additional simplifying assumption that the noncancer cause of death c_1 is also Weibull in form with common parameter γ, then

$$h_1(t) = \gamma\rho^{\gamma} t^{\gamma-1}$$

and the quantities of interest are

$$Q_2(\infty,x) = \int_0^{\infty} \gamma(\alpha + \beta x^m) t^{\gamma-1} e^{-(\alpha + \beta x^m + \rho^{\gamma}) t^{\gamma}}$$

$$= (\alpha + \beta x^m)/(\alpha + \beta x^m + \rho^{\gamma})$$

and

$$E = \int_0^\infty \gamma(\alpha + \beta x^m + \rho^\gamma) t^\gamma e^{-(\alpha + \beta x^m + \rho^\gamma) t^\gamma} \, dt$$

$$= \Gamma(1 + 1/\gamma) / (\alpha + \beta x^m + \rho^\gamma)^{1/\gamma} .$$

TABLE II-1

CALCULATION OF EXPECTED LIFETIME PROBABILITY OF DEATH DUE TO RESPIRATORY CANCER AND EXPECTED LIFE SPAN

			$h(t)$	$h_2(t)$					
Age	j	$t_j - t_{j-1}$ yrs.	$\alpha_j \times 10^5$	$\alpha_{2j} \times 10^5$	P_{1j}	P_{2j}	P_{3j}	Q_{2j}	E_j
<1	1	<1	1805.2	0.1	5.5396×10^{-5}	1.7890×10^{-2}	1.00000	9.9103×10^{-7}	0.99103
1-4	2	4	79.5	0.	0	3.1750×10^{-3}	.98211	0.	3.92221
5-9	3	5	41.5	0.	0	2.0729×10^{-3}	.97899	0.	4.88994
10-14	4	5	40.6	0.1	2.4631×10^{-3}	2.0279×10^{-3}	.97696	4.8798×10^{-6}	4.87983
15-19	5	5	111.9	0.1	8.9366×10^{-4}	5.5794×10^{-3}	.97498	4.8613×10^{-6}	4.86127
20-24	6	5	146.8	0.2	1.3624×10^{-3}	7.3131×10^{-3}	.96954	9.6599×10^{-6}	4.82994
25-29	7	5	143.5	0.5	3.4843×10^{-3}	7.1493×10^{-3}	.96261	2.3975×10^{-5}	4.79502
30-34	8	5	165.7	1.5	9.0525×10^{-3}	8.2510×10^{-3}	.95573	7.1371×10^{-5}	4.75809
35-39	9	5	235.1	6.5	2.7648×10^{-2}	1.1686×10^{-2}	.94784	3.0619×10^{-4}	4.71065
40-44	10	5	355.2	15.8	4.4482×10^{-2}	1.7603×10^{-2}	.93676	7.3339×10^{-4}	4.64169
45-49	11	5	563.3	37.1	6.5862×10^{-2}	2.7772×10^{-2}	.92027	1.6830×10^{-3}	4.53641
50-54	12	5	832.8	62.2	7.4688×10^{-2}	4.0785×10^{-2}	.89471	2.7250×10^{-3}	4.38098
55-59	13	5	1314.5	105.1	7.9954×10^{-2}	6.3612×10^{-2}	.85822	4.3642×10^{-3}	4.15245
60-64	14	5	1943.9	148.4	7.6341×10^{-2}	9.2621×10^{-2}	.80363	5.6814×10^{-3}	3.82842
65-69	15	5	2804.7	182.8	6.5176×10^{-2}	1.3085×10^{-1}	.72920	6.2176×10^{-3}	3.40131
70-74	16	5	4302.7	206.9	4.8087×10^{-2}	1.9357×10^{-1}	.63378	5.8982×10^{-3}	2.85076
75-79	17	5	6722.4	209.8	3.1209×10^{-2}	2.8546×10^{-1}	.51110	4.5527×10^{-3}	2.71001
80-84	18	5	9777.4	174.1	1.7806×10^{-2}	3.8668×10^{-1}	.36570	2.5142×10^{-3}	1.44408
85+	19	∞	17429.4	136.6	7.8373×10^{-3}	1.0000	.22398	1.7552×10^{-3}	1.28489
								$3.6547 \quad 10^{-2}$	71.32889
						Σ			

$Q_2(\infty) = 0.036,546,670$

$E^2 = 71.328,888$ years

REFERENCES

1. Chiang, C.L. 1968. Introduction to stochastic processes in biostatistics. John Wiley, New York.

2. Doll, R. 1971. The age distribution of cancer: Implications for model of carcinogenesis. J. Roy. Stat. Soc. A 134:133-166.

3. Gail, M. 1975. Measuring the benefit of reduced exposure to environmental carcinogens. J. Chron. Dis. 28:135-147.

4. U.S. Dept. of Health, Education and Welfare. 1973. Vital statistics of the United States, Vol. II--Mortality, Part A. Supt. of Documents, U.S. Government Printing Office. Stock No. 017-022-549-2.